DOCUMENT RESUME

ED 443 846 TM 031 491

AUTHOR Novak, John R.; Winters, Lynn; Flores, Eugene
TITLE Using Multiple Measures for Accountability Purposes: One District's Experience.
PUB DATE 2000-04-27
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Accountability; Compensatory Education; Elementary Secondary Education; Performance Based Assessment; Research Methodology; *School Districts; Student Evaluation; *Testing Programs
IDENTIFIERS *Long Beach Unified School District CA; *Multiple Measures Approach

ABSTRACT

This paper describes the approach to consolidating multiple measures of student achievement used by the Long Beach Unified School District (LBUSD) in the 1997-1998 reporting cycle. Beginning in 1996-1997, all California schools that served Title I students or were involved in the state's Coordinated Compliance Review process were required to submit a Student Achievement Report based on multiple measures of student achievement. The LBUSD was relatively well situated to deal with the challenges of this requirement because it had already implemented a district-wide testing program that used performance assessments at multiple grade levels. It was necessary to consider the challenges involved in using component weighting models for combining multiple measures. Problems in combining these measures lead the school district to implement a compensatory standards-based approach. Decision rules for this approach are described, and an example is provided of the decision matrix used to determine student proficiency. The final approach used by the LBUSD did not give the district any unfair advantage over other school districts, while at the same time the LBUSD was able to maintain an internal standard of performance that perhaps was higher than that set by the state. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 443 846

Using Multiple Measures for Accountability Purposes: One District's Experience

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

John R. Novak
jnovak@math.usc.edu

Lynn Winters
Lwinters@lbusd.k12.ca.us

Eugene Flores
Gflores@lbusd.k12.ca.us

Research, Planning, and Evaluation
Long Beach Unified School District
1515 Hughes Way
Long Beach, CA 98010

J.R. Novak

April 27, 2000

Paper presented at the 2000 American Educational Research Association Annual Meeting, New Orleans, LA.

TM031491

BEST COPY AVAILABLE

During the 1996-98 school years public school districts in the State of California were faced with data analysis and reporting challenges of unprecedented magnitude. In response to the requirements of the Improving America's Schools Act of 1994 (IASA), Title I, California local educational agencies (LEA's) were required to submit to the California Department of Education (CDE) reports of student achievement that were based on multiple measures. While there were some stipulations made by the California Department of Education regarding what measures were to be included and how those measures would be combined, for the most part school districts were left up to their own devices regarding the details of the process. This paper describes the approach to consolidating multiple measures that was used by the Long Beach Unified School District (LBUSD) in the 1997-98 reporting cycle.

Background

Beginning in the 1996-97 school year, all California schools that served Title I students, or were involved in the State of California's Coordinated Compliance Review (CCR) process, were required to submit to the State a school-level Student Achievement Report (SAR; see Appendix 1 for an example of a completed SAR) based on multiple measures of student achievement. This report summarized the percentages of students who were achieving at or above grade level standards, broken down by the following demographic categories: All Students, Specially-Funded Students (Title I and Migrant Education), Limited English Proficient (LEP) Students, Special Education Students, and Gifted and Talented (GATE) Students. . While there were no immediate sanctions, schools with less than 40% of their population meeting grade level standards (MGLS) would be identified as Program Improvement Schools. Schools receiving this designation would be subject to special scrutiny until they had been judged to be making adequate yearly progress towards the statewide goal of 90% of students meeting standards. Program Improvement schools would have to bear the stigma of being publicly identified as under-performing schools, and the downstream sanctions for continuing not to meet growth targets could be severe, up to and including reconstitution and a takeover by the State.

In the initial year of the SAR (1996-97) it was largely left up to the individual Districts to set their own performance standards. This latitude predictably resulted in a great deal of inconsistency in the rigor of local standards and in the quality of the assessments used to measure attainment of those standards. During the 1997-98 reporting cycle, the State decided to take a more directive approach. This task was made easier by the adoption of the Stanford Achievement Test, Ninth Edition (SAT9) as the official instrument of the Standardized Testing and Reporting (STAR) system and the linchpin of the new accountability system. For the first time since the demise of the California Learning Assessment System, all schools in California would be administering the same test to their students. A memo by Ruth McKenna from the office of the Superintendent of Public Instruction (April 15, 1998) issued the following guidelines for a multiple measures accountability system:

- At least one measure in each of reading/language arts and mathematics would be required in each grade from 1 through 12.

- At least two measures per subject area would be required in at least **one grade** in each of the grade spans of 3-5, 6-9, and 10-12.
- In grades 2-11 the SAT9 test would be used as one of the components of the accountability system.
- Grade-level performance on the SAT9 was designated as the 50th percentile. If class grades were used as a component, then a grade of C or better would constitute grade-level performance.

Challenges to school districts

Complying with these requirements proved to be very difficult for many school districts, especially for the 1996-97 reporting cycle. The guidelines for this reporting process were not released to districts until June of 1997, and reports were due to the State in November of 1997 incorporating student achievement for the previous year. Many smaller districts were not in the habit of collecting data that could be used for multiple measures decisions in a usable format. Prior to the 1996-97 school year all districts were required to administer standardized tests to their students and were required to report the results to the State. For the 1996-97 cycle the choice of which standardized test to administer was still left up to the individual district, and each district was required to include the results from that test as one of the multiple measures for the 1996-97 reporting cycle. However, not all districts were in the habit of ordering results from those tests in electronic formats, instead relying on hard-copy reports generated by the test publishers. This made the utilization of those scores for the purposes of State-level accountability difficult. Even more distressing for these districts was the fact that alternative assessments were either non-existent or not readily available. Districts without fully staffed research offices often had no other assessments to fall back on other than grades, and often (especially at the elementary level) those grades existed only in hard copy format, hand written by teachers on barely legible NCR copies. Just harvesting that data extended the capacities of these districts to the limit.

Long Beach Unified School District multiple measures strategy

The measures. LBUSD was relatively well situated to deal with these challenges even during the first year of implementation (1996-97). The District had already implemented a district-wide testing program utilizing performance assessments at multiple grade levels in Writing and Mathematics. The use of Reading Benchmarks was in the process of being phased in, and during the 1996-97 school year had been widely administered in grades 1-3. Most students in grades 2-6 had also taken tests on their Basic Math Facts (Addition, Subtraction, Multiplication and Division). District End-of-course exams were in the process of being developed for Mathematics courses at all grade levels. The large enrollment of the District could also support a well-staffed and well-equipped research staff to deal with data collection, consolidation, and reporting issues. Table 1 summarizes the measures that were utilized by the LBUSD for the purpose of making standards-based decisions during the 1997-98 reporting cycles.

Table 1 -- Measures Used to Determine Student Proficiency in 1997-98

	Reading-Language	Mathematics
Grade K	Benchmark books	None
Grade 1	Benchmark books	Grade 1 District Math Test
Grade 2	Benchmark books, SAT9 Reading Total	Math Facts, SAT9 Math Total
Grade 3	Benchmark books, Writing, SAT9 Reading Total	Math Facts, Open ended math, SAT9 Math Total
Grade 4	SAT9 Reading Total	Math Facts, SAT9 Math Total
Grade 5	Writing, SAT9 Reading Total	Math Facts, Open ended math, SAT9 Math Total
Grade 6	Writing, SAT9 Reading Total	Math Facts, Open ended math, SAT9 Math Total
Grade 7	SAT9 Reading total	Integer Test, SAT9 Math Total
Grade 8	Writing, SAT9 Reading Total	Integer test, open ended math, SAT9 Math Total
Grade 9	SAT9 Reading Total	SAT9 Math Total
Grade 10	Writing, SAT9 Reading Total	Open ended math, SAT9 Math Total
Grade 11	SAT9 Reading Total	SAT9 Math Total
Grade 12	Not included	Not included

Challenges of using component weighting models for combining multiple measures

Once multiple measures are available, districts are still faced with the task of how to go about making proficiency decisions based on multiple sources of information. Early approaches to this problem espoused by the CDE overly simplified the combination problem. In 1996-97 the following guideline was provided by the CDE:

- Districts will develop their own systems for weighting the multiple measures used to assess student performance. While there is no prescribed weight that should be assigned to each measure, districts should only include measures which contribute significantly to the overall performance assessment. If three measures are used, the minimum weight assigned to any one measure should be 25 percent. If two measures are used, the minimum weight assigned to any one measure should be 30 percent (CDE, 1997, p.5).

Several problems soon surfaced, not the least of them the widely varying scales and quality of the measures used. For example, in Table 1 we see that at Grade 3 decisions in Language Arts were based on 1) the highest level attained on the Reading Benchmarks, which is an ordinal variable tied in to grade level in a descriptive fashion; 2) the score on the District Writing Assessment, which is rubric scored on a scale from 1-6; and 3) the national percentile rank (NPR) on Total Reading from the SAT9 test, ranging from 1-99. Extensive attention has been paid to the reliability of the SAT9 test by the publisher, but there are still questions about the validity of this instrument as the ultimate arbiter of the quality of a student's school experience. The Benchmarks and the Writing Assessment may better reflect the valued outcomes of the District, but there are serious questions about the reliability of those measures. How to combine such disparate measures into a single decision is a question for which there is no clear answer. Rudner (2000), and

others, have examined approaches to weighting component scores to create a composite score with maximal reliability or maximal criterion validity. Ryan and Hess (1997) applied a Discriminant Analysis approach to the very problem of combining multiple measures for purposes of Title I reporting. While these approaches are certainly defensible, no clear-cut choice has emerged. There is the additional consideration that even if a suitable approach were chosen, as the methodology becomes more complex and circuitous, the likelihood that school districts with limited expertise and limited resources will be able implement it decreases.

Missing Data. A much larger problem is that of missing data. The guidelines from the CDE required that decisions about proficiency be made on a student-by-student basis. The component weighting approaches all involve creating linear combinations of scores, setting a cutpoint for proficiency, and then applying that function to individual student's scores to determine whether or not that student had attained proficiency. In a perfect world, all students would have complete data, but those who work in school districts know that the world is far from perfect. Given that out of three measures, the minimum weight assigned to any measure must be at least 25%, any student missing any of the three measures is almost certain to have a composite score below the cutpoint for proficiency. A general principle that may be applied here is that *the more measures that you are using in the assessment process, the larger the proportion of students who will be missing one or more of those measures.*

As an example of the magnitude of the missing data problem, consider Table 2. This table provides information about students enrolled in Grades 3, 5, 6, and 8 in the 1997-98 school year. These were the grade levels where LBUSD was using three measures to assess Mathematics proficiency. You can see that the percentages of students missing at least one of the measure ranged from 12.1% in the 5th grade to 32.6% in the 8th grade. One of the reasons for the magnitude of the missing data problem is the ad hoc nature of the State's accountability system. Note that the most problematic measure is the Math Facts, and that is largely due to the fact that teachers and schools were not made aware early enough about the potential use of that measure for accountability purposes. Hence their investment in applying that assessment tool was not as great as it would have been had there been more time to inform them about the role of this measure in this high-stakes accountability system.

Table 2 – Counts and percents of students missing components of the Mathematics assessment

Grade	Total	Missing SAT9	Missing OEM	Missing MF	Missing at least one measure
3	7180	217 3.0%	371 5.2%	900 12.5%	1251 17.4%
5	6704	191 2.8%	361 5.4%	423 6.3%	809 12.1%
6	6473	300 4.6%	354 5.5%	640 9.9%	1070 16.5%
8	6336	506 8.0%	487 7.7%	1482 23.4%	2068 32.6%

SAT9 – Stanford 9 Total Math National Percentile Rank
OEM – Score on District Open-ended Mathematics Assessment (1-6 scale)
MF -- # of Basic Math Facts Tests Passed (Maximum of 4)

The missing data situation was even more serious during the 1996-97 reporting cycle, and it was immediately apparent that component weighting approaches were not going to be feasible. Examination of the Federal guidelines for Title I assessment provided justification for alternative approaches. According to those guidelines,

- The decision of whether to combine scores to produce school of district results, and the method used to combine scores, should be based on producing reliable and valid information for the purpose(s) and use(s) intended. If school/district results are reported for each assessment separately, overall judgements of performance in a content area can be based on the pattern of results, using either a conjunctive approach (requiring a particular level of performance on each assessment) or a compensatory approach (allowing performance on the various assessments to counterbalance each other) (OESE, 1997, p. 44).

The Federal guidelines acknowledged the fact that academic proficiency is far from being a unidimensional construct, and that "A single measure or approach is unlikely to adequately measure the knowledge, skills, and complex procedures covered by rigorous content standards. (OESE, 1997, p. 42)." It was also apparent that due to the large amount of missing data, component weighting methodologies would be inadequate for providing "reliable and valid information" for the purpose of accountability at the District/school levels. Armed with this knowledge, LBUSD decided to explore alternatives to component weighting schemes that would better serve the intent of the IASA guidelines.

An alternative approach to combining multiple measure

Many psychometricians working on issues of combining multiple measures have a tendency to focus on the numerical/technical aspects of the task. For the most part their conceptions of reliability and validity are guided by classical test theory and focus on correlational information. School districts, however, are much less interested in the technical details and are much more interested in the consequences. Traditional approaches to validity have focused in on content, construct, and criterion-related validity, but have often neglected the consequential basis that must underlie any valid accountability process (Messick, 1989). The highly public nature and high stakes of the California school accountability system ensures that school districts will never lose this focus.

The nature of the accountability beast precludes the application of a conjunctive approach to combining multiple assessments. Such an approach is diagrammed in Table 3 using two hypothetical measures. In this type of approach, in order to be judged proficient a student much achieve proficiency in <u>all</u> of the measures used. While such a standard is certainly admirable, and could certainly be used at the classroom level within a school or district to set the bar higher and promote achievement, to do so in a high-stakes, public accountability context would be politically suicidal. The more hurdles that are placed before a student, the more likely that they will stumble over one of them. Using a conjunctive approach for high stakes accountability purposes, whether at an institutional level (i.e., identification of underperforming schools) or at the individual level (as in High

School graduation requirements) is probably not advisable, and would be tantamount to standing beside each of the hurdles waiting to shoot any runners who stumble.

Table 3 – Example of a conjunctive decision rule

	Measure 1	
Measure 2	Proficient	Not Proficient
Proficient	Proficient	Not Proficient
Not Proficient	Not Proficient	Not Proficient

Accordingly, for the 1996-97 reporting cycle LBUSD decided to implement a compensatory approach to combining multiple measures. In this type of system, low (or missing) scores on one component can be compensated for by high scores on other components. In addition to being politically sensible, such an approach also makes it relatively easy to accommodate students who are missing components of the system. From a humanistic perspective, instead of each of the measures being viewed as an obstacle to success, they are transformed into opportunities to succeed. Decision matrices implementing this approach were developed at each grade level, SAR reports were generated, and the reports were submitted to the CDE with a rationale for our approach.

Not only did the LB approach to combining multiple measures prove to be acceptable to the State, it was actually adopted by the State as one of the recommended approaches to combining multiple measures for the 1997-98 reporting cycle (CDE, 1998). That year was the first year that the SAT9 test was administered on a statewide basis, and in the guidelines that instrument played a featured role in several ways:

- The SAT9 had to be included as one of the components in Language Arts and Mathematics in grades 2-11.
- The cutpoint for proficiency on the SAT9 would be the 50th National Percentile. This reflects the statewide goal of having 90% of the population achieving at or above the 50th percentile—truly a goal almost worthy of the mythical Lake Wobegon.
- If a compensatory approach were used, no student scoring below the 30th percentile on the SAT9 could be judged as being proficient, regardless of performance on other measures.

Implementing a compensatory approach

The basic framework. The approach used by LBUSD to implement a compensatory standards-based accountability system based on multiple measures will be illustrated for a case utilizing two measures--the SAT9 Total Math national percentile rank, and the score on the District's Open-ended Mathematics assessment (OEM). Decision rules for this system are captured in a contingency table, and the first step was to appropriately categorize the data so the skeleton for the table could be constructed. The OEM was scored on a rubric, and valid scores could range from 1 to 6. Students who participated in the assessment but whose papers could not be scored either due to lack of response or an off-topic response were assigned zeros. Some students (736 at the 10th grade level) had

not participated in the assessment and so had missing scores. It was decided that for this measure there was already an adequate balance between the continuous (more score points) and the discrete (fewer score points) so no further collapsing of scores was done. The SAT9 scores range from 1 to 99, and so it was decided to collapse the valid scores into six performance bands. Those bands were from 1-29, 30-39, 40-49, 50-59, 60-69, and 70-99. The 30th and 50th percentiles provided obvious cutpoints due to their salience in the guidelines provided by the state, and an additional cutpoint was added at the 40th percentile to create the opportunity to make finer distinctions. Cutpoints above the 50th percentile were added in the interest of added precision and of symmetry. Table 4 contains a contingency table based on these decisions.

Table 4 -- Framework for a grade-level standards decision matrix using two measures.

		SAT9 Math Total Percentile Rank					
Open-ended Math	**Missing**	**1-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-99**
Missing							
0							
1							
2							
3							
4							
5							
6							

Decision rules. The next step was to determine for each cell in the matrix whether or not students with that combination of scores would or would not meet grade-level standards (MGLS). These decisions were informed both by the expectations imposed by the State that grade-level performance equates to the 50th percentile on the SAT9, and by local standards for the other measures created by the District. Within the District, the standard for proficiency on the OEM was a score of 4 or more. This information made it easy to fill in the cells in the 3x3 submatrix in the lower right quadrant of Table 4. Those students would have attained proficiency on both the SAT9 and the District assessment, and so they would be judged as MGLS. Note that if a conjunctive approach were utilized, these would be the only students judged as meeting standards. Using a compensatory approach, though, the District can make decisions such as allowing a student who only achieved a score of 3 on the OEM to be judged to be MGLS if that student had scored in the 60-69 SAT9 performance band. Here the excess achievement on the SAT9 provides evidence that the student did not work up to potential on the OEM.

In the original formulation of the District's policy the decision matrix looked very much like Table 5. In this table the notation 'MGLS' indicates that students in that cell have met grade-level standards. Several features are notable about this table. First, notice the triangular shape in the lower right of the table, incorporating the compensatory aspect. You can see this, for example, in the cell corresponding to an OEM score of 4 and a SAT9 score in the 40-49 range. Students in this range did not meet the state requirement for proficiency but did satisfy the District standard. Here the tie goes to the District. This can be justified on two counts. First, because we can. That is, if the State provides the leeway to make a decision that will benefit the District, then the District would be

foolish not to take advantage of it. Another more objective reason is that there is considerable variation in the difficulty of the OEM from year-to-year. This score is based on the students' response to a single prompt, and such a score is very sensitive to variation in prompt difficulty and other unintended characteristics of the prompt. If we err, however, it tends to be on the side of making the OEM more difficult than intended, and so on the average the OEM score will tend to underestimate the true ability of the student.

Table 5 -- Original version of the decision matrix

	SAT9 Math Total Percentile Rank						
Open-ended Math	**Missing**	**1-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-99**
Missing					MGLS	MGLS	MGLS
0							
1							MGLS
2						MGLS	MGLS
3					MGLS	MGLS	MGLS
4	MGLS			MGLS	MGLS	MGLS	MGLS
5	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
6	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS

The final features to note about Table 5 are the cells in the right part of the top row and the bottom part of the first column. These correspond to students who are missing either the OEM (top row) or the SAT9 (first column). Note that these students are judged as proficient or not based solely on the measure that they did have a score for.

Table 6—Final version of the decision matrix

	SAT9 Math Total Percentile Rank						
Open-ended Math	**Missing**	**1-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-99**
Missing					MGLS	MGLS	MGLS
0					MGLS	MGLS	MGLS
1					MGLS	MGLS	MGLS
2					MGLS	MGLS	MGLS
3					MGLS	MGLS	MGLS
4	MGLS			MGLS	MGLS	MGLS	MGLS
5	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
6	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS

Modifications to the rules. As more information trickled in from the CDE about the nature of the accountability system, especially with respect to potential adverse consequences of being identified as a Program Improvement school, as well as the very public nature of the process (results for all schools posted on the CDE website to the accompaniment of great fanfare), the resolve of the District to stick to the high road on setting standards wavered. Thus the final version of the decision matrix differed somewhat from the original and is presented in Table 6. Note that an additional six cells have been added to the MGLS category, and that any student scoring above the 50th percentile is automatically judged proficient regardless of their score on the OEM. Again, the driving force behind this decision is the desire to not "shoot ourselves in the

foot", but that decision can also be supported on firm rational grounds based on the uncertain reliability and validity of the District assessment.

This approach can easily be extended to three measures, and an example of a decision matrix for three measures is provided in Appendix 2. Theoretically, this process could be extended for an indefinite number of measures, but the amount of work required to create and implement the decision matrix increases multiplicatively. If more than three measures are used it might be worthwhile to explore other options for making decisions, such as some type of profile analysis. Missing data, however, might prove problematic for these approaches.

Table 7—Counts of students in each cell for Grade 10 Mathematics assessments

	SAT9 Math Total Percentile Rank						
Open-ended Math	**Missing**	**1-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-99**	**Total**
Missing	215	336	65	31	29	25	34	735
0	6	84	12	4	4	2	3	115
1	74	779	218	90	71	33	19	1284
2	58	503	241	129	142	106	55	1234
3	15	181	122	92	111	101	72	694
4	4	55	48	56	58	87	106	414
5	4	23	29	11	44	51	96	258
6	6	12	16	28	67	115	438	682
Total	382	1945	742	432	517	511	794	5416

Is the LBUSD approach valid?

We will briefly discuss this question relative to the alternatives of component weighting and conjunctive approaches, informed by the data in Table 7. For the purposes of this accountability process, validity is measured by the degree to which correct decisions are made about students' proficiency. We do not want students to be judged as not proficient only because they are missing one (or more) of the components of the assessment, or because of the lack of reliability of homegrown assessments. First, compared to the component weighting method, the compensatory approach classifies a total of 102 students, or 1.9% (unless otherwise noted percents will be based on the total population size) of the population, as proficient despite missing one of the measures. These students would not meet the cutpoint for proficiency using a component weighting approach.

Second, note that there are a total of 807 students (14.9%) who scored above the 50th percentile on the SAT9, but did not attain a score of 4 on the District's own OEM. This is indicative of the rigorous local standards that the LBUSD has set for performance. Using a conjunctive model, however, the District would be penalized for setting high standards, whereas our modification to the compensatory model allows the District to maintain both a high internal level of performance without paying a price in the accountability arena. Finally, note that only 140 students (2.5%) achieving scores lower than the 50th percentile on the SAT9 were judged as proficient using the compensatory model; of those students, 95 were in the 40-49 band on the SAT9 and thus were largely

within the margin of error on the SAT9 score. These factors indicate that the LBUSD did not gain any unfair advantage from utilizing this system, and at the same time was able to maintain an internal standard of performance that was perhaps higher than that set by the state without being unduly penalized.

Conclusion

The high priority and very public nature of the State level accountability process in California ensure that these issues will not go away soon. During the 1998-99 accountability cycle the State of California gave districts even less latitude to set local standards and reverted to using the SAT9 test as the sole assessment instrument. According to the legislation authorizing the Standardized Testing and Reporting (STAR) Program, the State is to move towards utilizing multiple measures for accountability purposes as those measures are developed and are shown to be valid and reliable. During the 1998-99 cycle the California Standards portion of the STAR test (known as the STAR augmentation) was piloted but the results were not incorporated into the system. The augmentation will be administered again during the current cycle, but, true to the ad hoc nature of accountability in California, school districts are currently administering tests to their students while they do not even know if or how the results of those tests will be used to judge them.

References

CDE (1997). *Guidelines for identifying IASA, Title I, Program Improvement Schools during the period of transitional assessment*. Draft document from the Office of the Superintendent of Public Instruction, California Department of Education, May 9, 1997.

CDE (1998). *Multiple Measures: Models for combining measures to determine whether students meet grade-level standards*. Memo prepared by the School and District Accountability Division, California Department of Education, April, 1998

Jang, Younghee (1998). *Implementing standards-based multiple measures for IASA, Title I accountability using TerraNova Multiple Assessment*. Paper presented at the 1998 American Educational Research Association Annual Meeting, San Diego, CA, April, 1998.

McKenna, R. (1998). *Standards-based accountability*. Memo from the Office of the Superintendent of Public Instruction , California Department of Education, April 15, 1998.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), In *Educational Measurement, Third Edition*, (pp. 13-104). New York: MacMillan

OESE (1997). *Guidance on standards, assessment, and accountability*. Draft document from the Office of Elementary and Secondary Education, Compensatory Education Programs, U. S. Department of Education, March 2, 1997.

Rudner, L. (2000). *Informed test component weighting*. Unpublished manuscript, Maryland Assessment Research Center for Education Success & ERIC Clearinghouse on Assessment and Evaluation.

Ryan, J. M. & Hess, R. K. (1997). *Weighting assessment indicators to maximize the accuracy of students' Title I classifications: An overview of technical issues*. Paper presented at the 27th Annual National Conference on Large Scale Assessment, Colorado Springs, CO, June, 1997.

1997-98 Student Achievement School Report

California Department of Education — **Consolidated Applicatio**

Purpose: This form is used to report standards-based achievement data for each school in the district by program. Please submit one copy of this form per school. Duplicate forms as necessary. (See revised instructions for 1997-98)

Agency: Long Beach Unified
CD Code: 1964725

School: School Name Here — **School Code: School Code**

Consider this site for: ☐ CA Distinguished School Program ☐ Blue Ribbon School ☐ IASA Title I Achieving School

Submission: ☐ Original ☐ Page not applicable ☒ Revison **Date: 03/04/99**

Line	Student Category	Enrollment	Reading/Language Arts			Mathematics			Average Percent
			Number Evaluated	Meeting or Exceeding Standards		Number Evaluated	Meeting or Exceeding Standards		
				Number	Percent		Number	Percent	
A	All Students	1097	1097	672	61.3%	998	631	63.2%	62%
B	Specially-Funded Students								
1	Title I/SCE - Targeted Assistance								
2	Migrant Education	2	2	0	0.0%	2	1	50.0%	25%
3	Limited English Proficient								
3a	R-FEP	133	133	73	54.9%	133	84	63.2%	59%
3b	LEP-Content	127	127	30	23.6%	118	37	31.4%	27%
3c	LEP-ELD	127	127	30	23.6%	118	37	31.4%	27%
4	Special Education	129							
4a	Group 1	129	129	31	24.0%	117	26	22.2%	23%
4b	Group 2								
5	Gifted and Talented	304	304	275	90.5%	304	287	94.4%	92%

Appendix 1: Sample School Achievement Report

Appendix 2: Decision Matrix for Three Measures

Long Beach Unified School District
Grade Level Standards Decision Matrix for Grades 3, 5, 6, 8 Mathematics

Final Version

		SAT-9 Percentile Rank						
Math Facts	**Performance**	**Missing**	**1-29**	**30-39**	**40-49**	**50-59**	**60-69**	**70-99**
Missing	Missing					MGLS	MGLS	MGLS
	0					MGLS	MGLS	MGLS
	1					MGLS	MGLS	MGLS
	2					MGLS	MGLS	MGLS
	3					MGLS	MGLS	MGLS
	4	MGLS			MGLS	MGLS	MGLS	MGLS
	5	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
	6	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
Passed 0 Tests	Missing					MGLS	MGLS	MGLS
	0					MGLS	MGLS	MGLS
	1					MGLS	MGLS	MGLS
	2					MGLS	MGLS	MGLS
	3					MGLS	MGLS	MGLS
	4	MGLS				MGLS	MGLS	MGLS
	5	MGLS				MGLS	MGLS	MGLS
	6	MGLS			MGLS	MGLS	MGLS	MGLS
Passed 1 Test	Missing					MGLS	MGLS	MGLS
	0					MGLS	MGLS	MGLS
	1					MGLS	MGLS	MGLS
	2					MGLS	MGLS	MGLS
	3					MGLS	MGLS	MGLS
	4	MGLS				MGLS	MGLS	MGLS
	5	MGLS			MGLS	MGLS	MGLS	MGLS
	6	MGLS			MGLS	MGLS	MGLS	MGLS
Passed 2 Tests	Missing					MGLS	MGLS	MGLS
	0					MGLS	MGLS	MGLS
	1					MGLS	MGLS	MGLS
	2					MGLS	MGLS	MGLS
	3					MGLS	MGLS	MGLS
	4	MGLS				MGLS	MGLS	MGLS
	5	MGLS			MGLS	MGLS	MGLS	MGLS
	6	MGLS			MGLS	MGLS	MGLS	MGLS
Passed 3 Tests	Missing					MGLS	MGLS	MGLS
	0					MGLS	MGLS	MGLS
	1					MGLS	MGLS	MGLS
	2					MGLS	MGLS	MGLS
	3	MGLS			MGLS	MGLS	MGLS	MGLS
	4	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
	5	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
	6	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
Passed 4 Tests	Missing					MGLS	MGLS	MGLS
	0					MGLS	MGLS	MGLS
	1					MGLS	MGLS	MGLS
	2					MGLS	MGLS	MGLS
	3	MGLS			MGLS	MGLS	MGLS	MGLS
	4	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
	5	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS
	6	MGLS		MGLS	MGLS	MGLS	MGLS	MGLS